

# Visualization

ALCF: Getting Started Workshop 2011

Mark Hereld & Joe Insley

# Our Goals

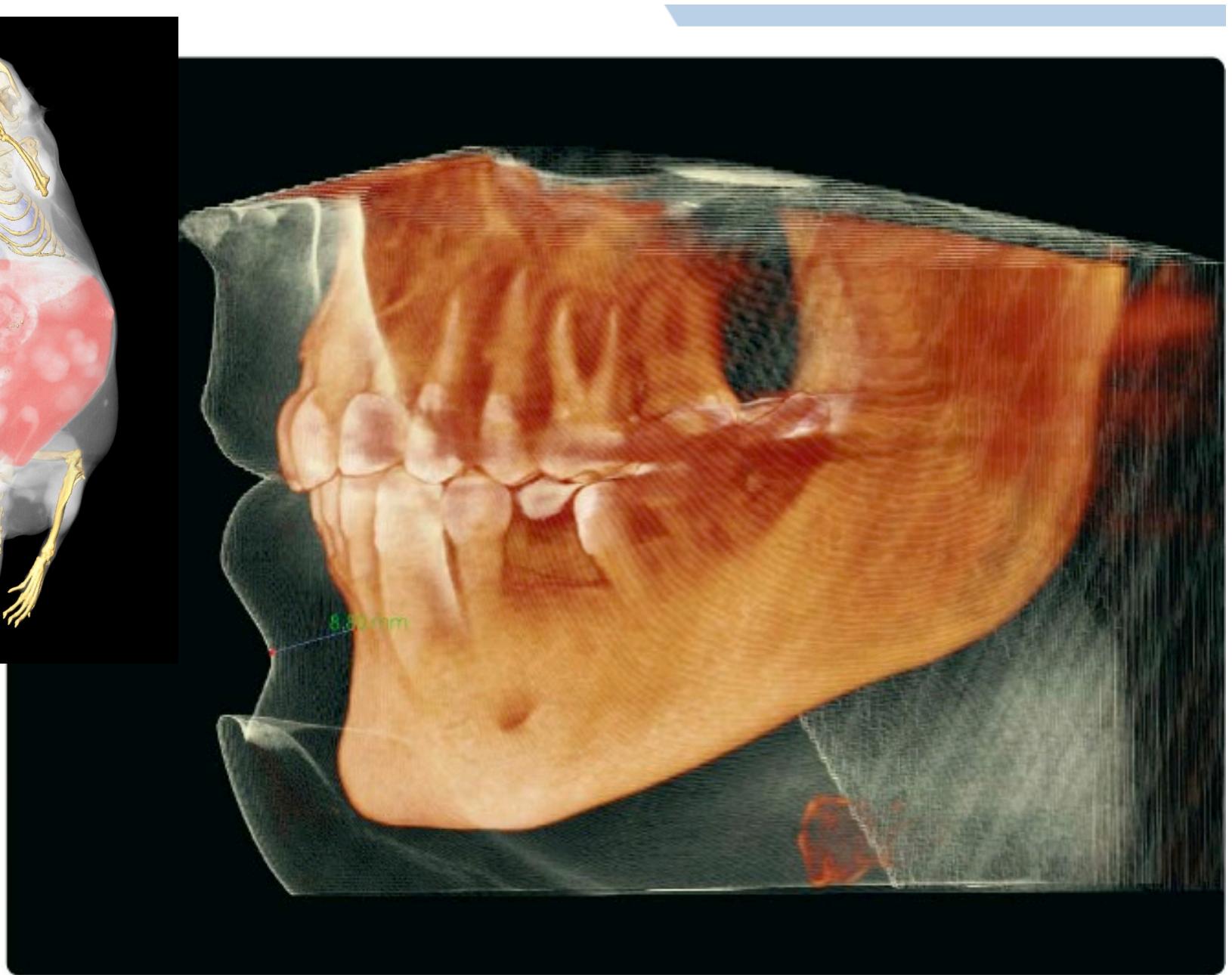
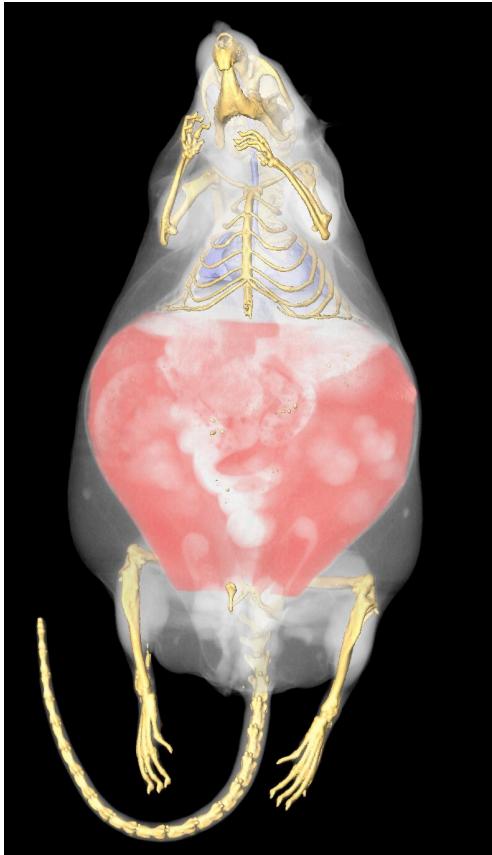
- Introduce you to Eureka
  - Examples of visualizations
  - Visualization tools available to you
  - Your large datasets
  - Things to think about
- 
- Practical tutorial
    - Getting started
    - Grokking ParaView
    - Visualization basics

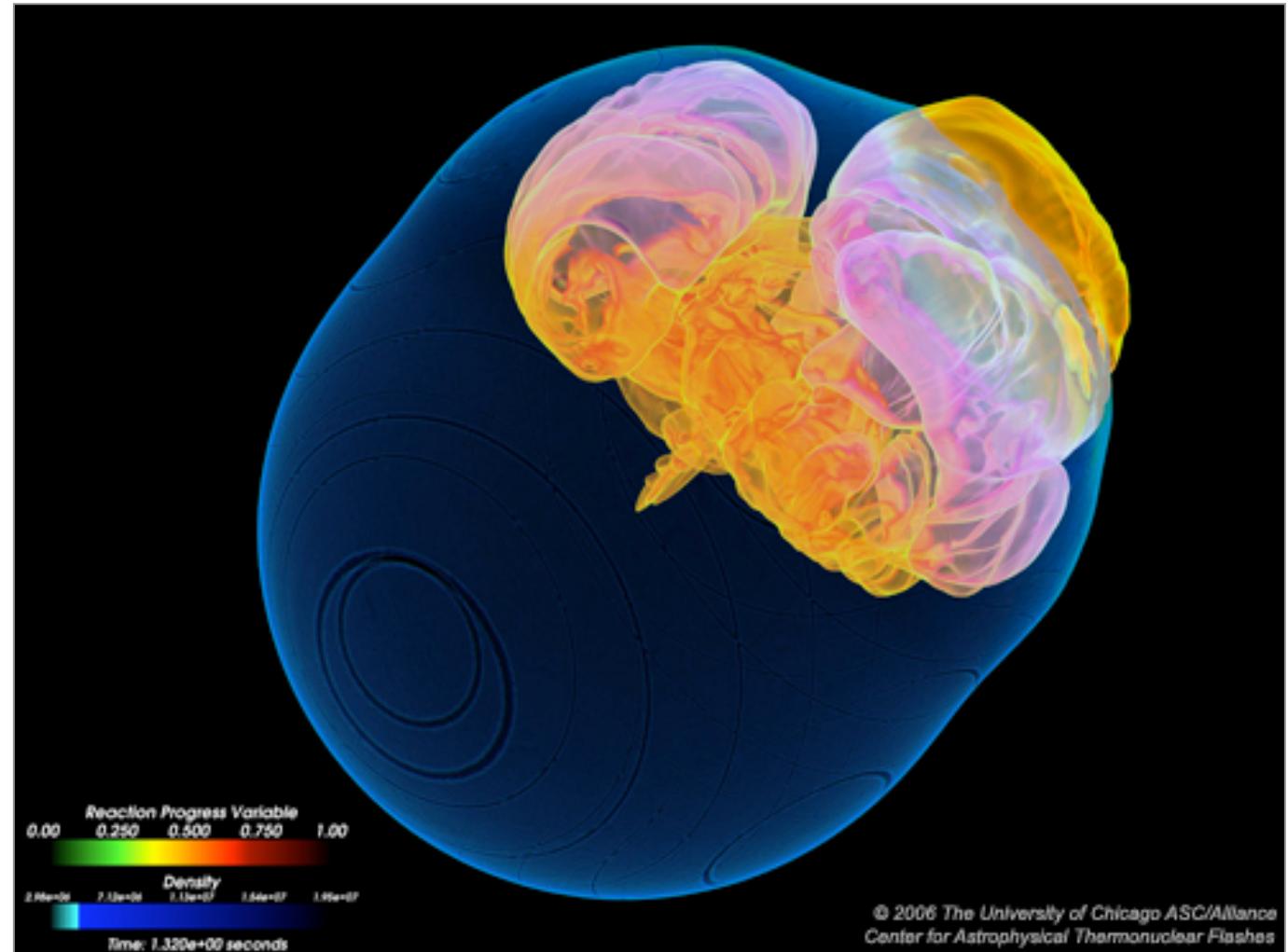


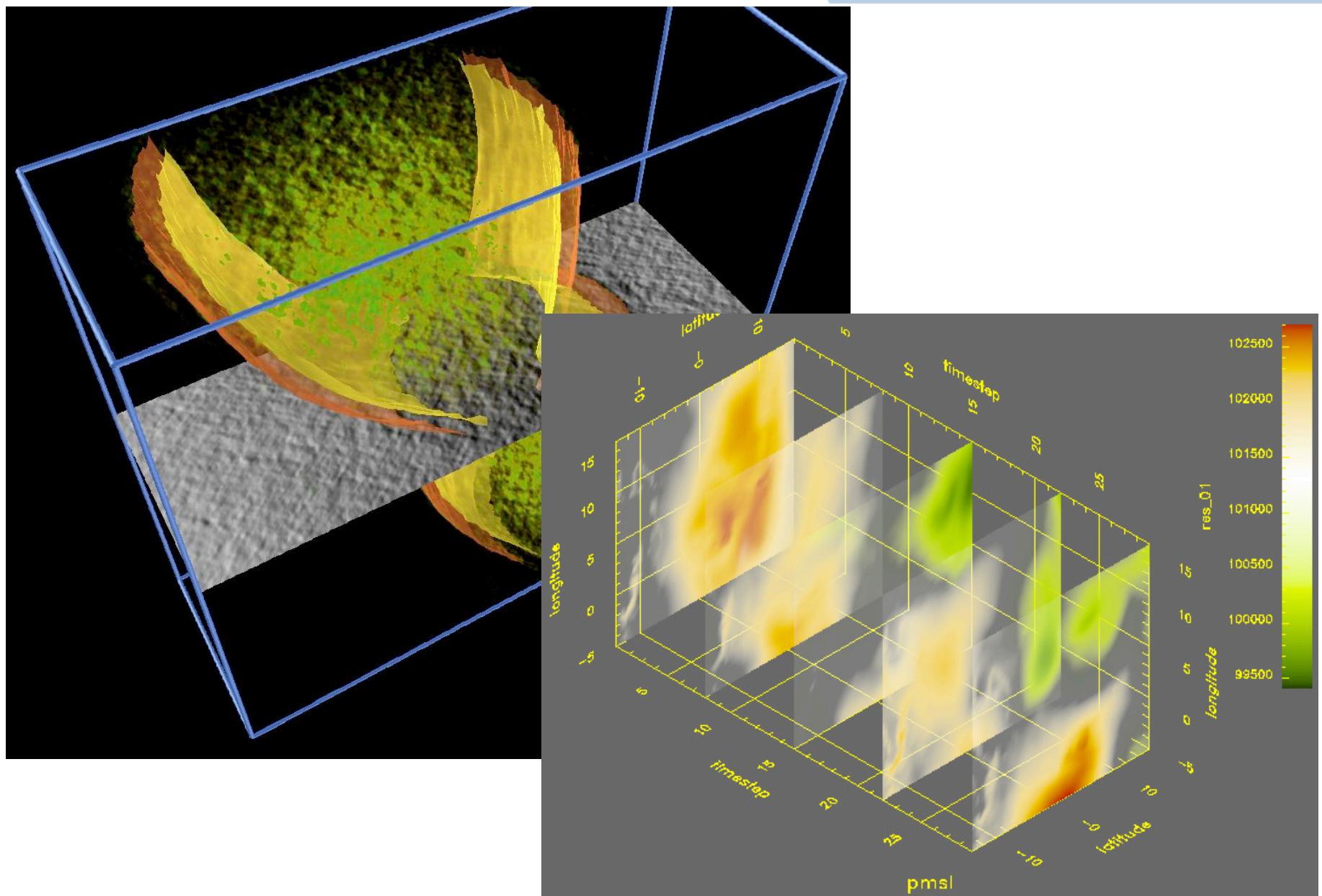
# ALCF's Eureka and Gadzooks

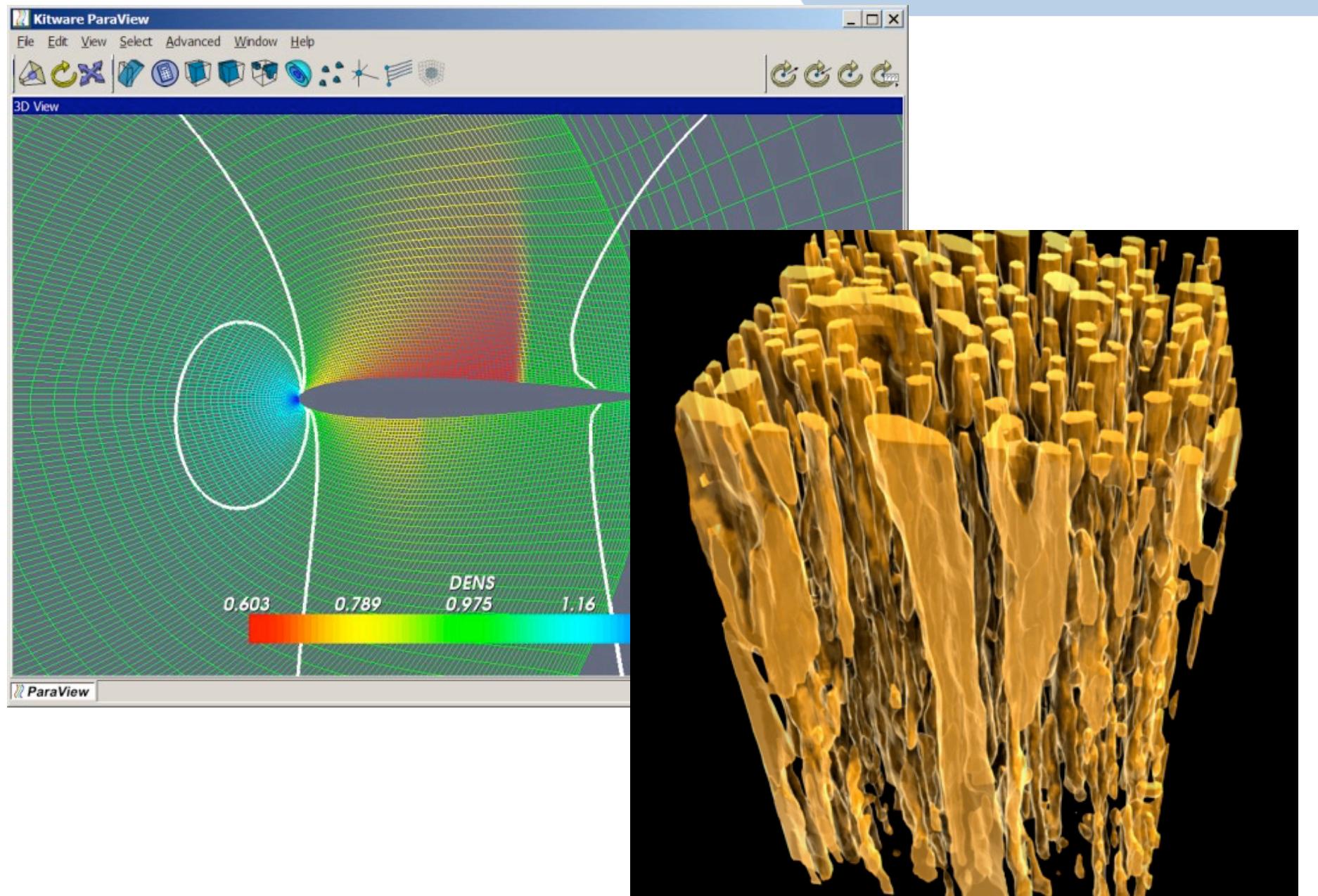
- 100 + 4 dual quad core servers
- 200 + 8 Quadro FX5600 graphics engines
- 312 Gbytes of total frame buffer RAM
- 3.2 TB of total system RAM
- Each node (server):
  - Dual quad core CPU
  - 2 GPUs
  - 1.5 GB frame buffer RAM
  - 32 GB system RAM

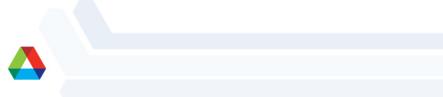
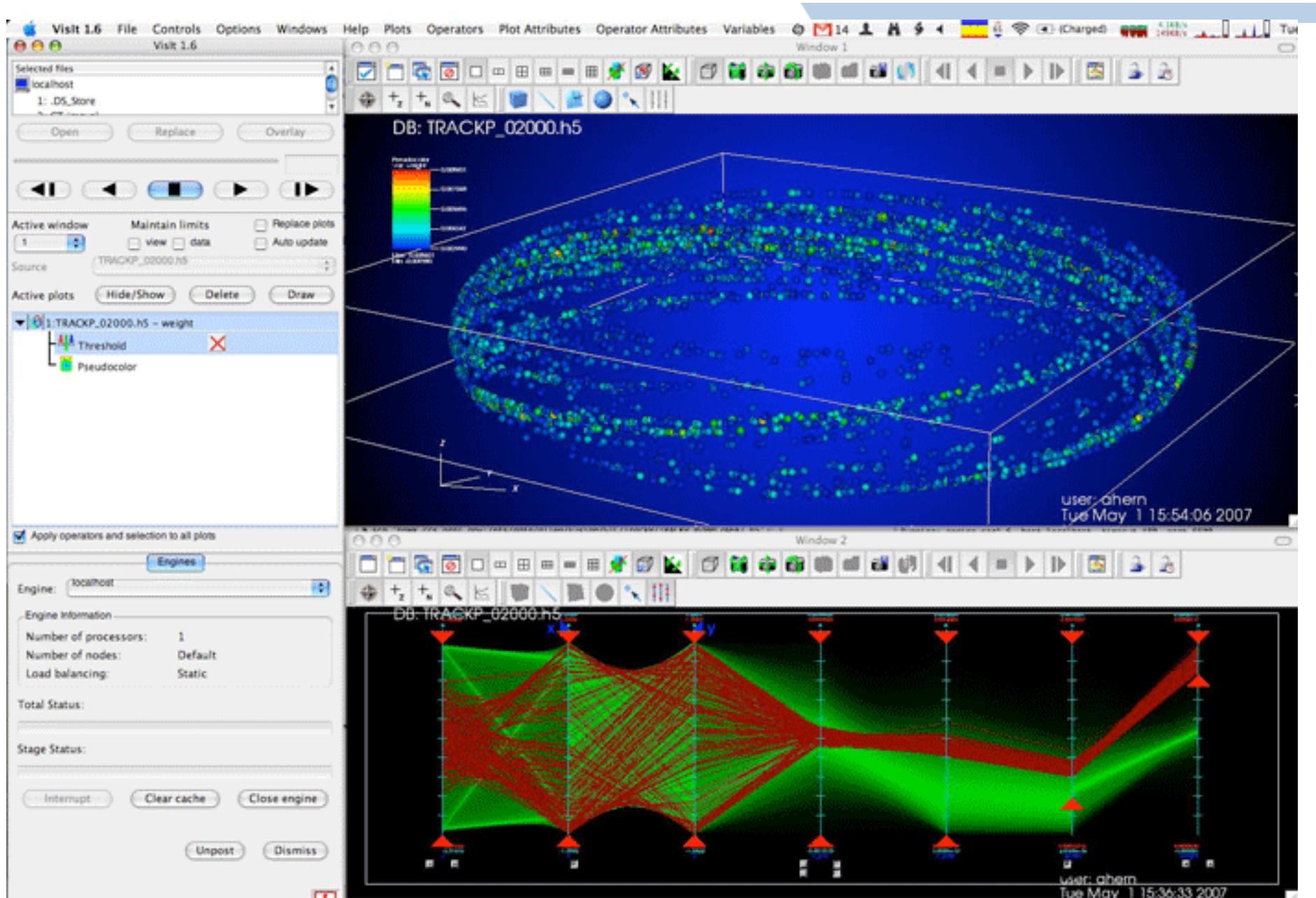










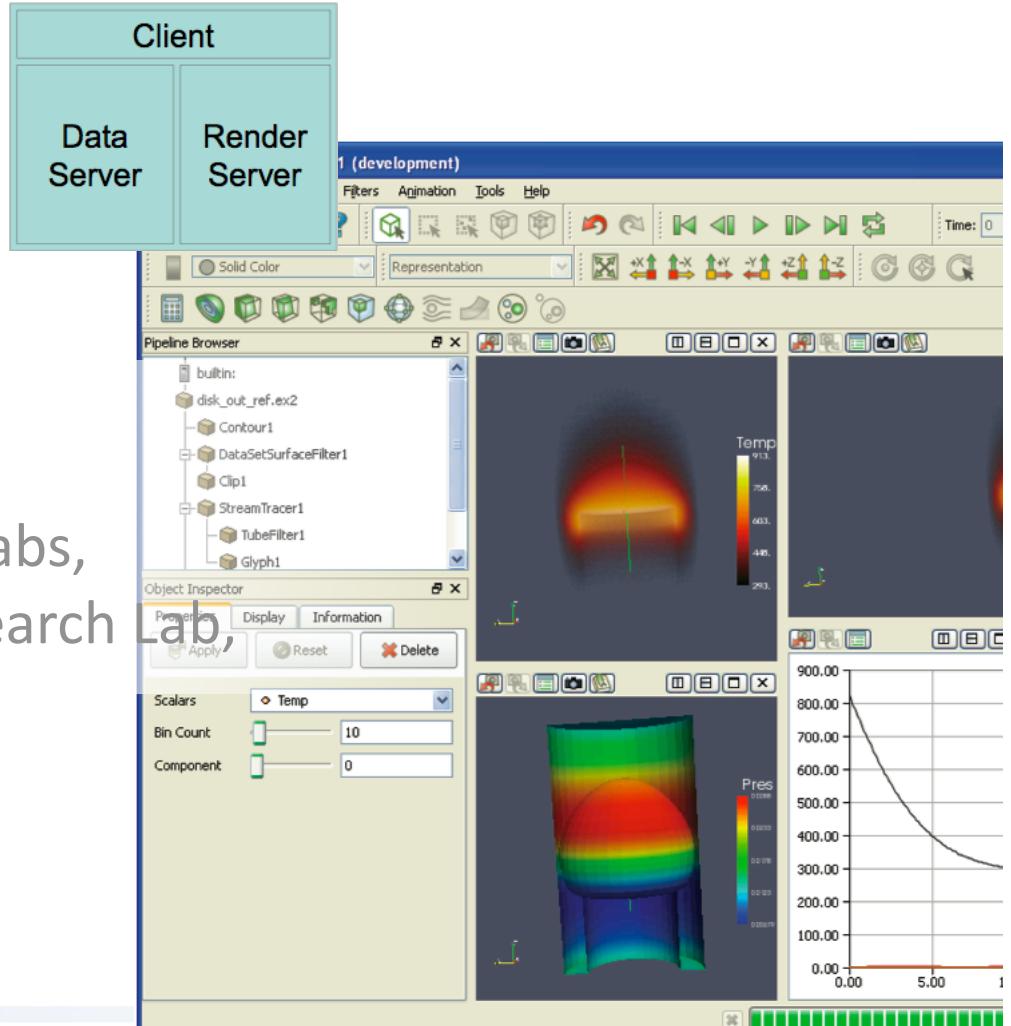
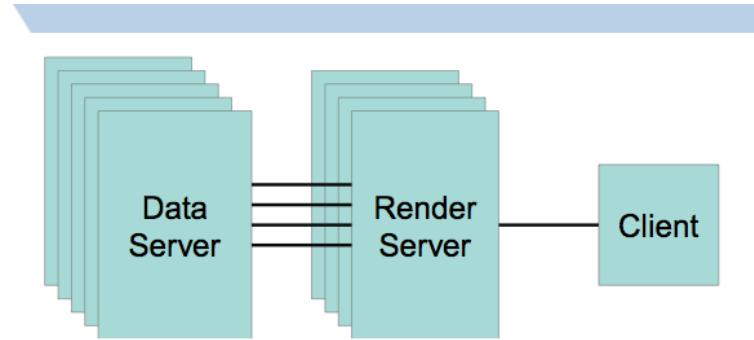


# All Sorts of Tools

- Visualization Applications
  - VisIt
  - ParaView
  - EnSight
- Domain Specific
  - PyMol, RasMol
- APIs
  - VTK: visualization
  - ITK: segmentat & registration
- GPU performance
  - vI3: shader-based vol ren
  - Scout: GPGPU acceleration
- Analysis Environments
  - Matlab
  - Parallel R (ORNL)
- Utilities
  - GnuPlot
  - ImageMagick
- Visualization Workflow
  - VisTrails

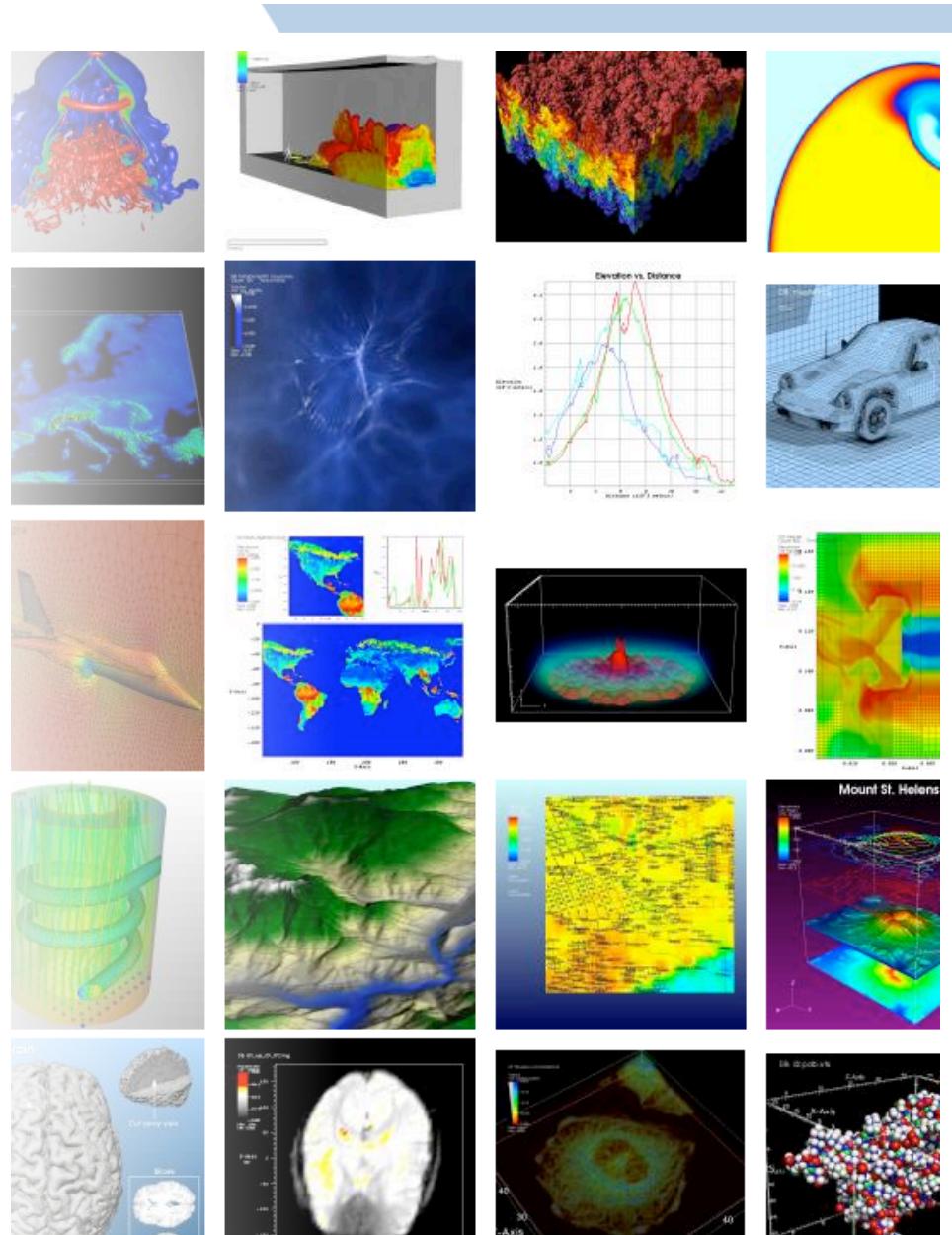
# ParaView Overview

- Parallel Visualization Application
- Open source
- VTK + Tcl
- Python scripting
- Interactive and batch
- About
  - Kitware, Sandia National Labs,  
CSimSoft, LANL, Army Research Lab,  
...and community
  - <http://www.paraview.org>



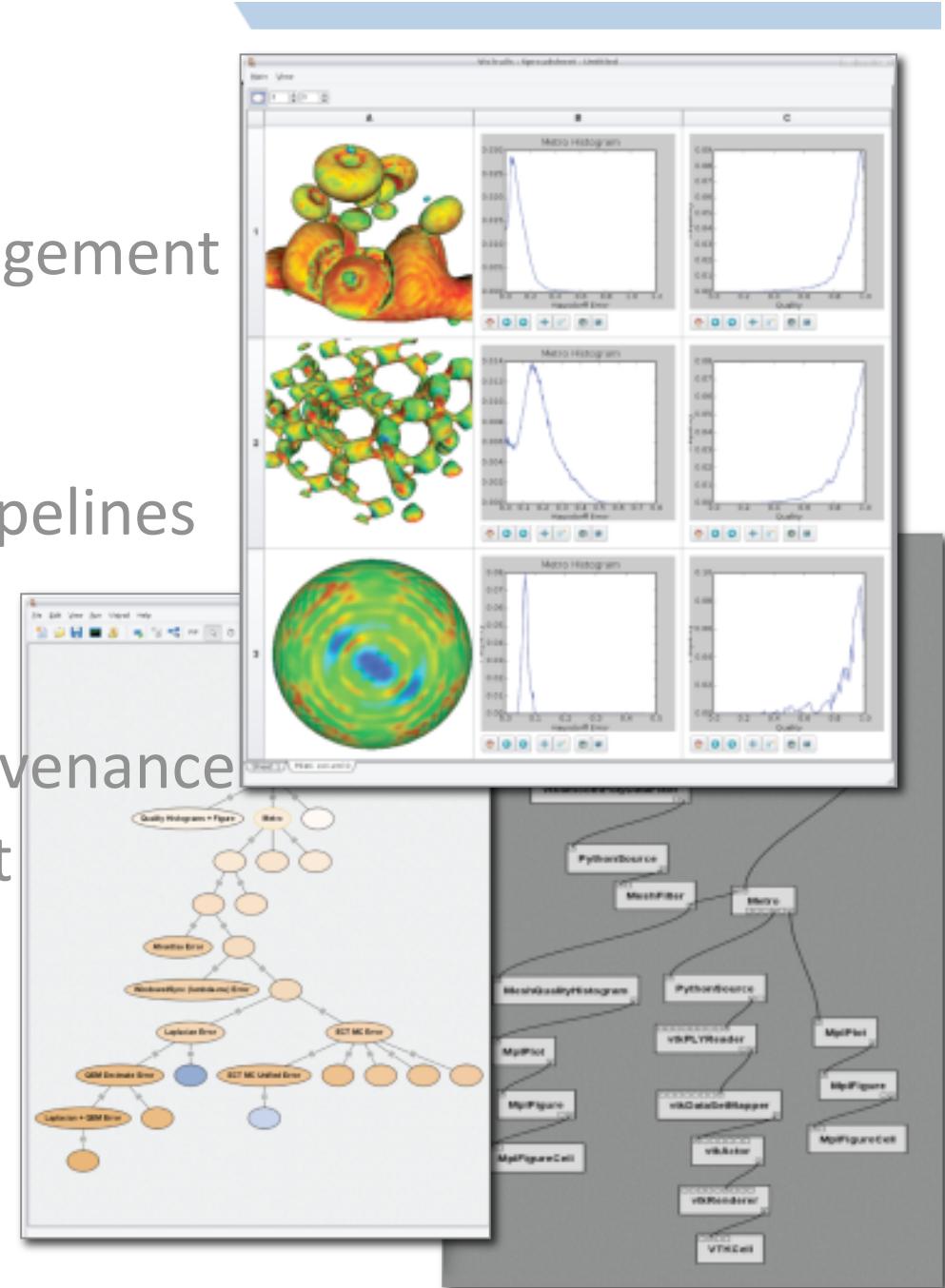
# VisIt Overview

- Parallel interactive visualization application
- About
  - DOE ASCI
  - <https://www.llnl.gov/visit>



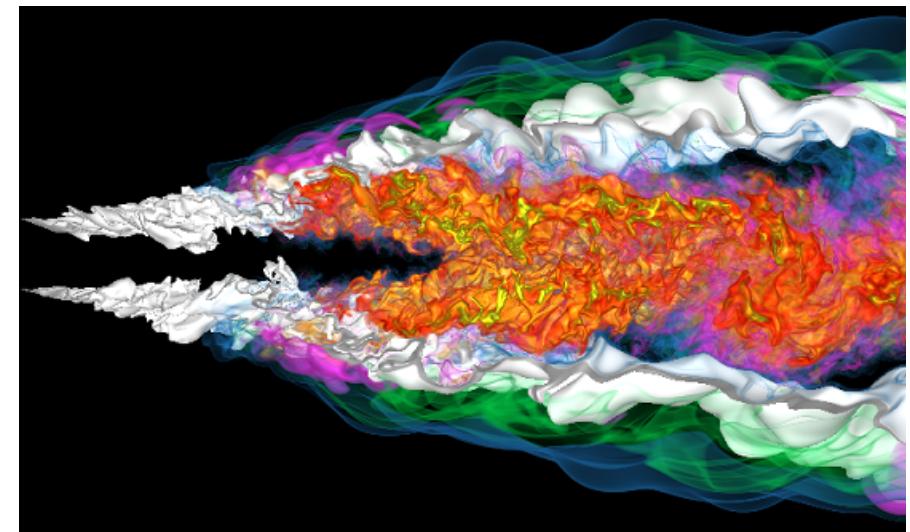
# VisTrails

- Scientific workflow management for visual data analysis
- Construct and execute pipelines
  - Visual programming
  - VTK, ITK, and Matplotlib
- History tree captures provenance
- Visualization spreadsheet
- About
  - <http://www.vistrails.org>



# In situ analysis and data reduction

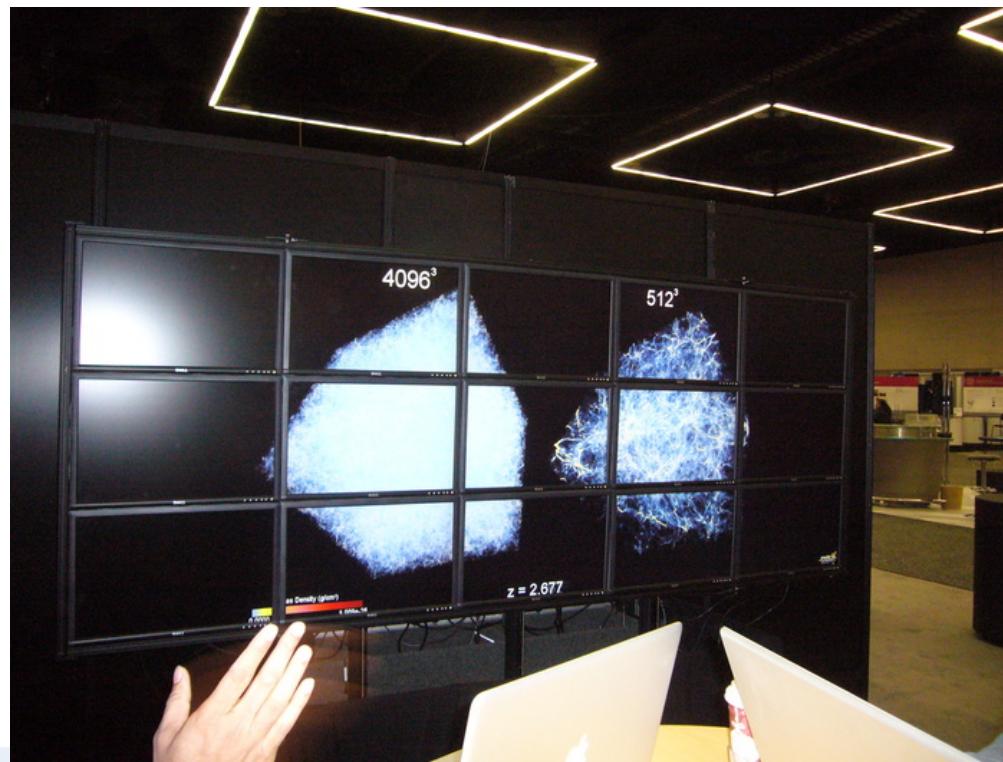
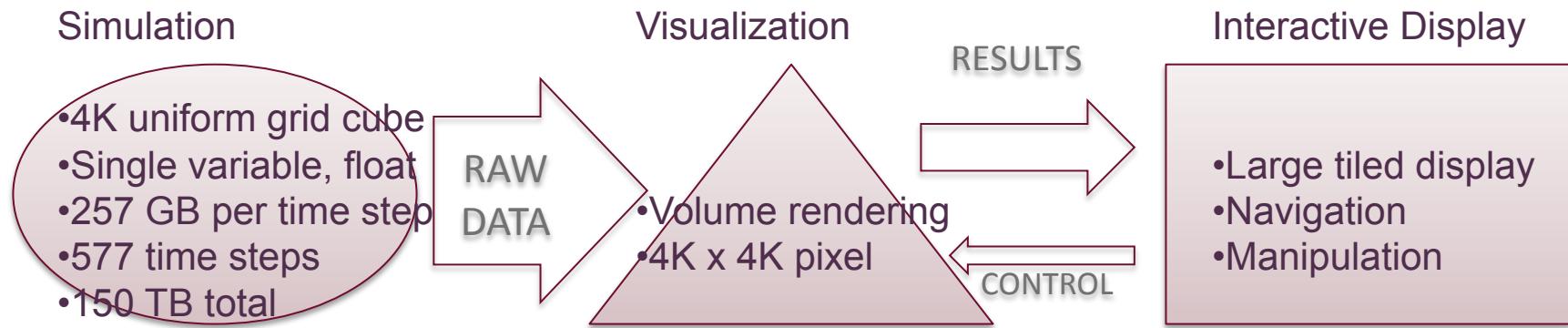
- Incorporate analysis routines into the simulation code
  - operate on data while it is still in memory
- Potential for significant reduction the I/O demands
  - application scientist identifies features of interest
  - compress data of less interest



C. Wang, H. Yu, and K.-L. Ma, "Application-driven compression for visualizing large-scale time-varying volume data", IEEE Computer Graphics and Applications, accepted for publication.

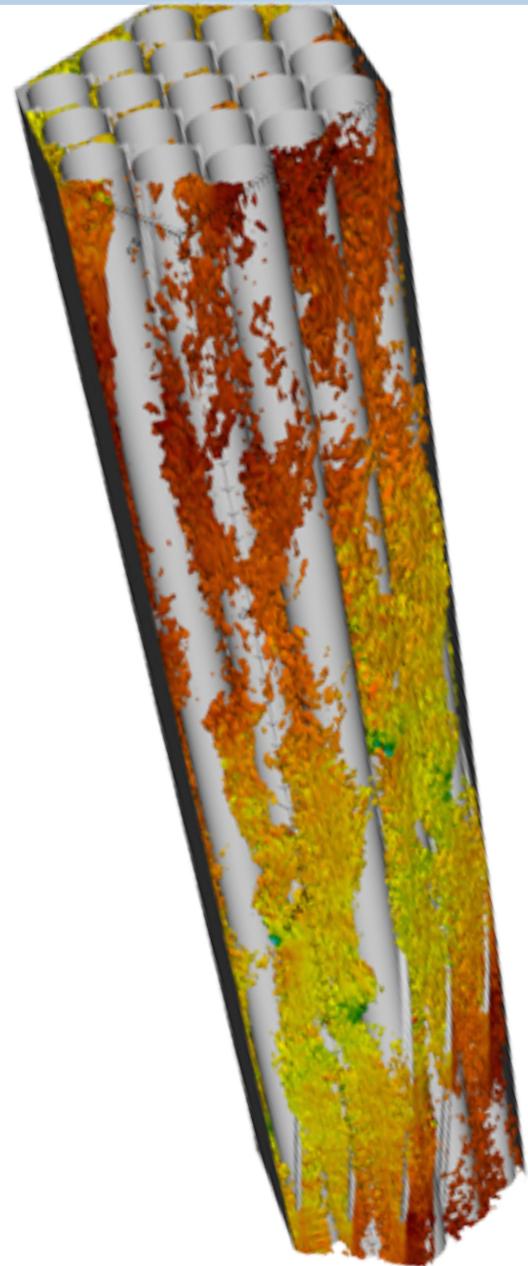


# Wide Area Experiments



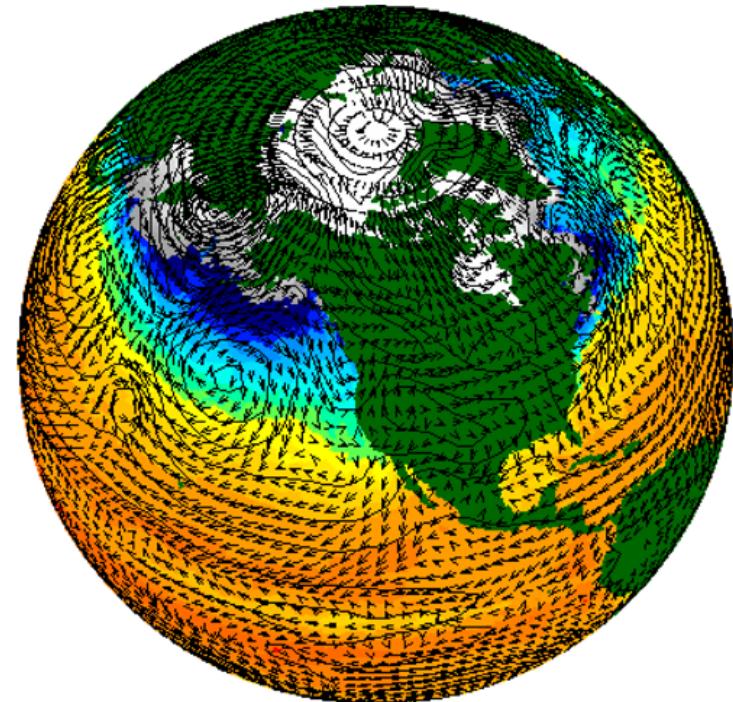
# Nuclear Reactor Simulation

- Preliminary studies
  - 4.5 million elements
  - 7 variables per element
  - 20 K timesteps
  - Total data produced 2.5 TB
- Science runs
  - 3 – 4 runs with 120 million elements
  - Several runs at  $\frac{1}{2}$  and  $\frac{1}{4}$  resolution
  - 90 K timesteps
  - Total data produced 900TB – 1.2 PB



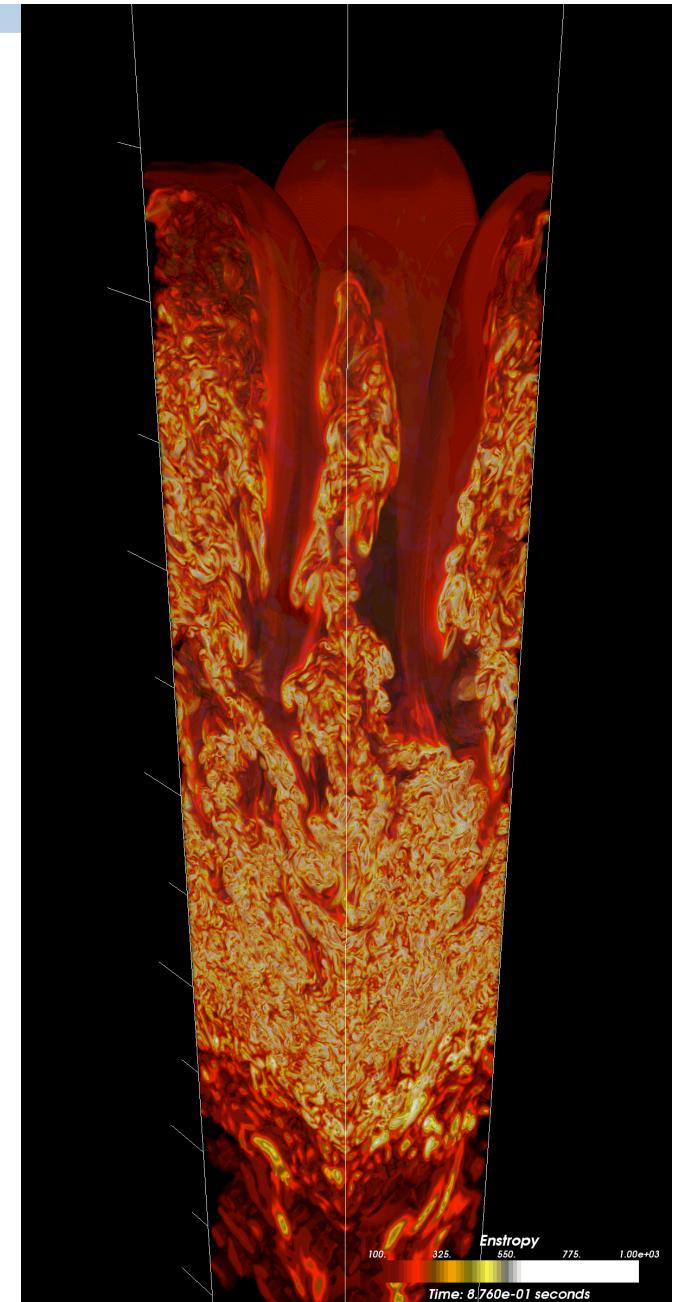
# Climate Modeling

- Preliminary studies
  - 50-100 with 3 million grid points (1 M atmosphere, 2 M ocean)
  - 100 variables per grid point (30 vectors, 70 scalars)
  - Simulating 5 - 10 years of climate
  - Total data produced 30 -124 TB
- Science runs
  - 50 runs with 6 million grid points
  - Simulating 100 years of climate
  - Total data produced 1.2 PB



# Astrophysics

- Preliminary studies
  - ~80 with 67 M grid points
  - ~5 with 536 M grid points
  - 6 variables (1 vector, 3 scalars)
  - ~1800 time steps
  - Total data produced 78 TB
- Science run\*
  - 10242 x 4096 grid points
  - 6 variables (1 vector, 3 scalars)
  - ~1800 time steps
  - Total data produced 48 TB



— \*3-5 times bigger allocation is needed

# Your Goals

- What do you look to get out of analysis process?
- What analysis tools are you currently using?
  - What are the limitations?
- Do you do real-time exploration or batch processing?
  - What is the role of real-time exploration?
- Batch?
  - Percentage of your analysis time spent in either mode?
- Do you look at images, movies or graphs?
  - What is role of each (e.g. graphs for science, images publications, movies for talks)



# Your Footprint

- How long do your simulations run?
  - Is the result a time series?
  - How many files does that produce?
- How much of your simulation time is I/O?
- What are your dataset sizes?
  - checkpoint files, variables, species, analysis files
- How long do you spend on analysis, what is the fraction of compute versus human?

